Towards Real-World Test-Time Adaptation: Tri-Net Self-Training with Balanced Normalization

Yongyi Su
South China University of Technology
eesuyongyi@mail.scut.edu.cn

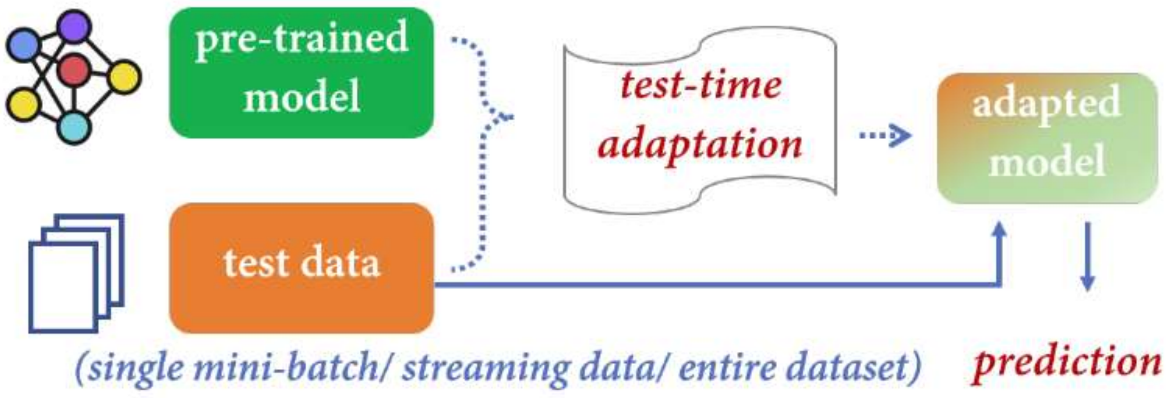
Xun Xu
Institute for Infocomm Research, A*STAR
alex.xun.xu@gmail.com

Kui Jia
South China University of Technology
kuijia@scut.edu.cn

AAAI 2024

Background

Test Time Adaptation



- stationary class distribution
- static domain shift

Motivation

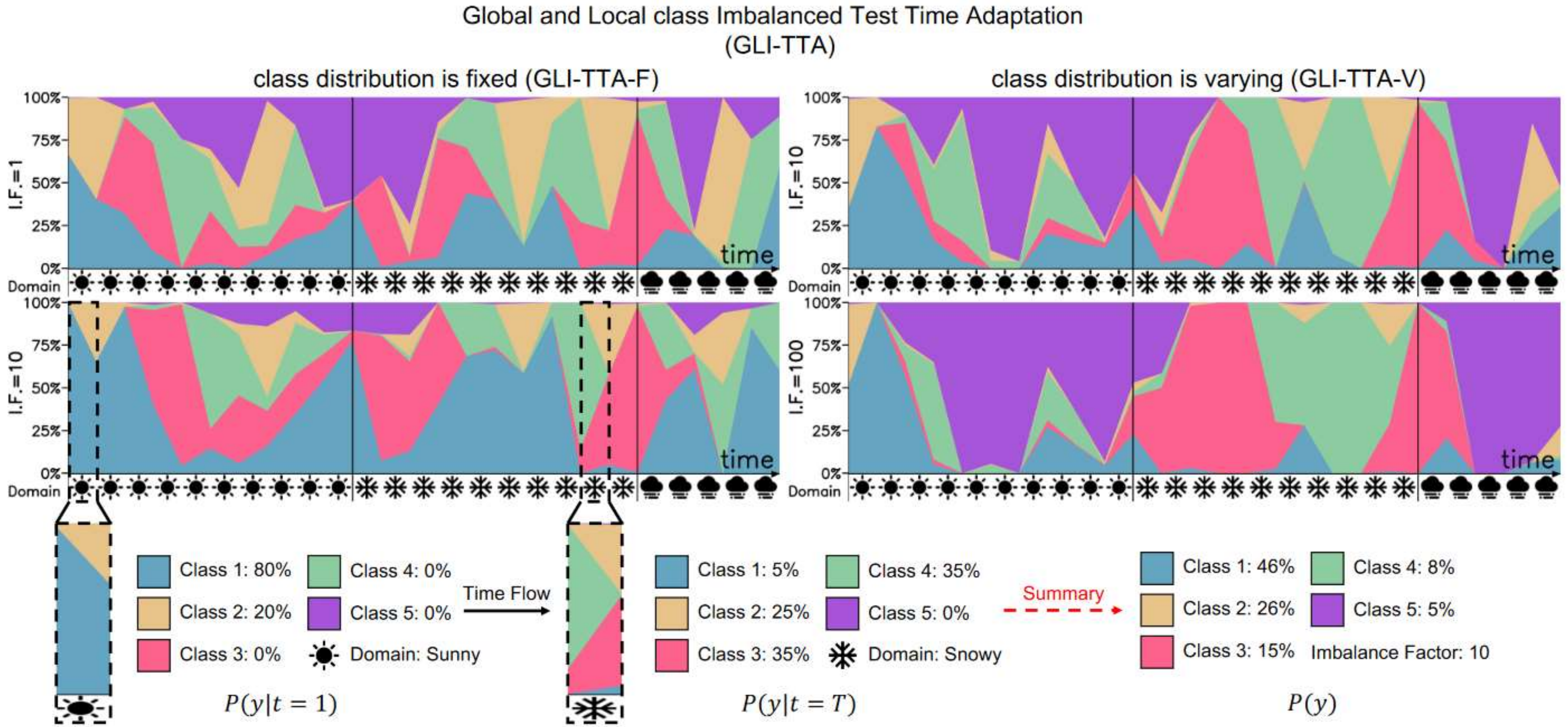


Figure 1: Illustration of two challenging real-world TTA scenarios. Different colors indicate the proportions of semantic classes, horizontal axis indicates testing data domain (e.g. different corruptions) may shift over time and different imbalance factor ($I.F.$) controls the degree of global imbalance. We expect the testing data stream to exhibit both local and global class imbalance, termed as "class distribution is fixed (**GLI-TTA-F**)" and this distribution may also evolve over time, termed as "class distribution is varying (**GLI-TTA-V**)".

Methods

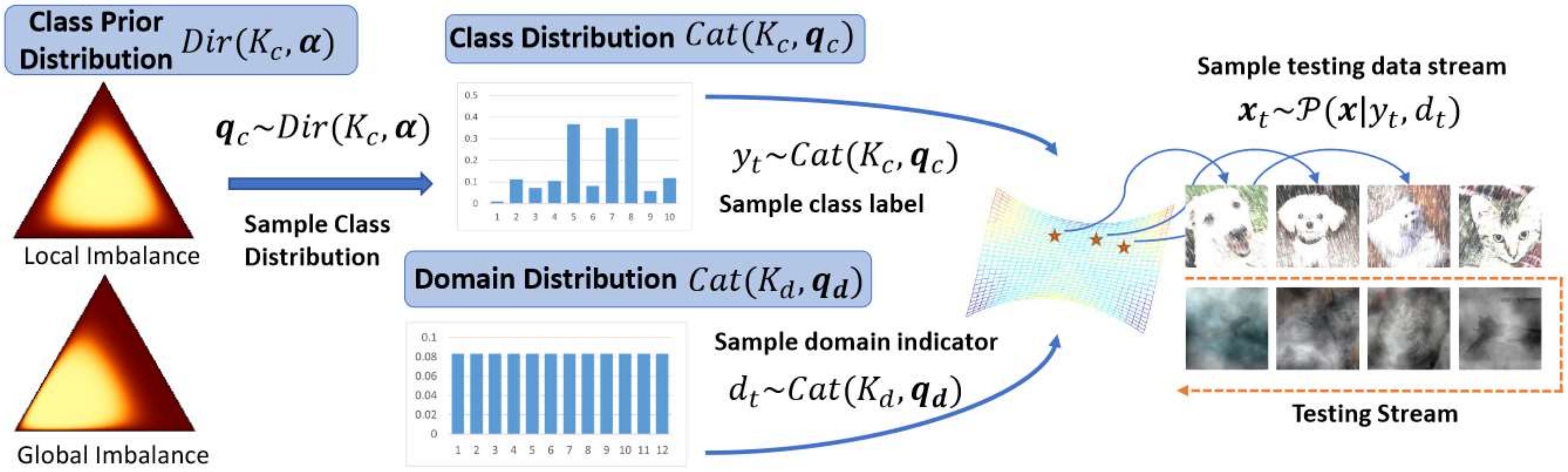


Figure 2: An illustration of the proposed real-world TTA simulation protocol with a hierarchical probabilistic model. A non-uniform α results in globally imbalanced testing data distribution.

Methods

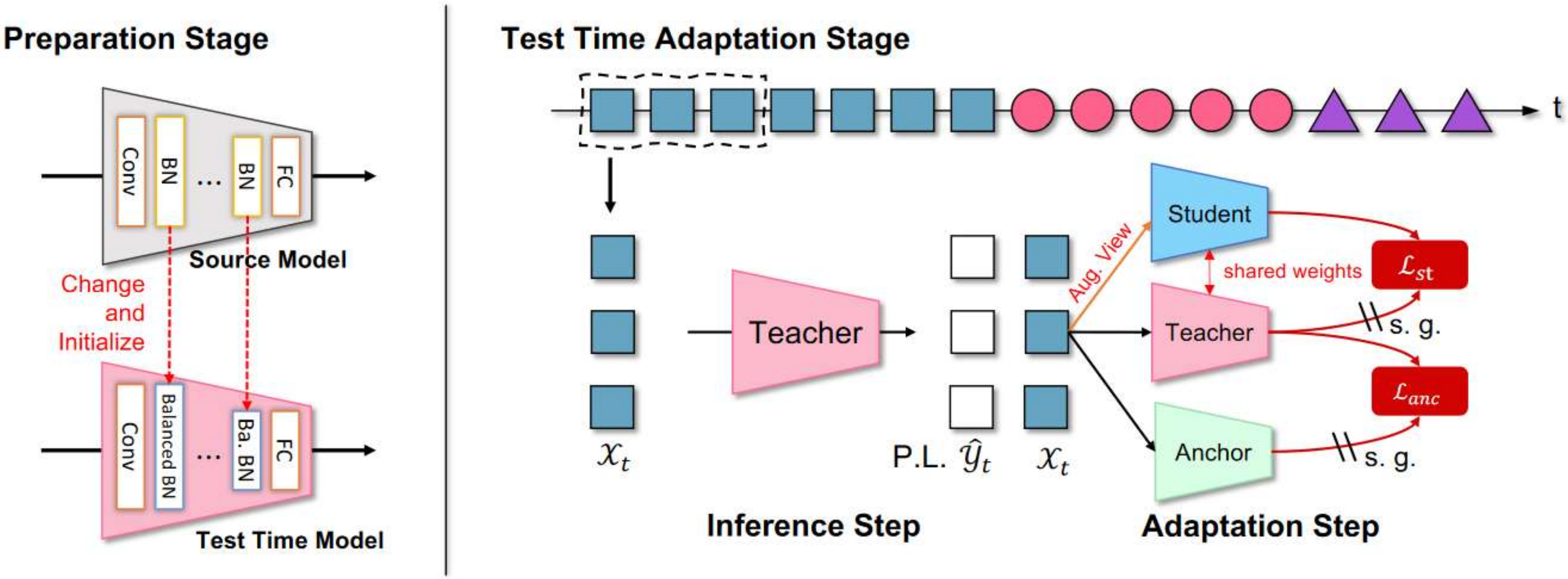


Figure 3: Illustration of the proposed method. We replace the Batchnorm layer of the source model with our proposed Balanced Batchnorm for imbalanced testing set. During test time adaptation, we optimize the combination of self-training loss $\mathcal{L}_{st}$ and anchor loss $\mathcal{L}_{anc}$.

Balanced Batch Normalization

$$\mu_k^t = (1-\eta)*\mu_k^{t-1} + \eta * \mu_k^t = \mu_k^{t-1} + \eta*(\mu_k^t - \mu_k^{t-1}) = \mu_k^{t-1} + \eta*\delta_k$$

$$\mu_k^t = \mu_k^{t-1} + \delta_k, \quad \sigma_k^{2\,t} = \sigma_k^{2\,t-1} - \delta_k^2 + \eta \sum_{b=1}^{B} \mathbb{1}(\hat{y}_b = k) \frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} \left[(F_{bhw} - \mu_k^{t-1})^2 - \sigma_k^{2\,t-1} \right]$$

$$s.t. \quad \delta_k = \eta \sum_{b=1}^{B} \mathbb{1}(\hat{y}_b = k) \frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} (F_{bhw} - \mu_k^{t-1}), \tag{1}$$

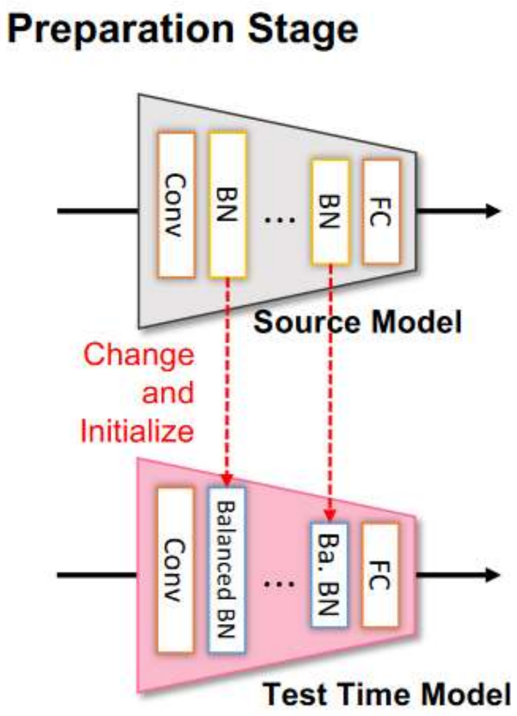
$$\mu_g = \frac{1}{K_c} \sum_{k=1}^{K_c} \mu_k^t, \quad \sigma_g^2 = \frac{1}{K_c} \sum_{k=1}^{K_c} \left[\sigma_k^{2\,t} + (\mu_g - \mu_k^t)^2 \right]. \tag{2}$$

$$\mu_k^t = \mu_k^{t-1} + (1-\gamma)\delta_k + \gamma \frac{1}{K_c} \sum_{k'=1}^{K_c} \delta_{k'},$$

$$\mu_k^t = (1-\eta)*\mu_k^{t-1} + \eta*(\gamma*\mu_k^t + (1-\gamma)*\frac{1}{K}\sum \mu_k^t)$$

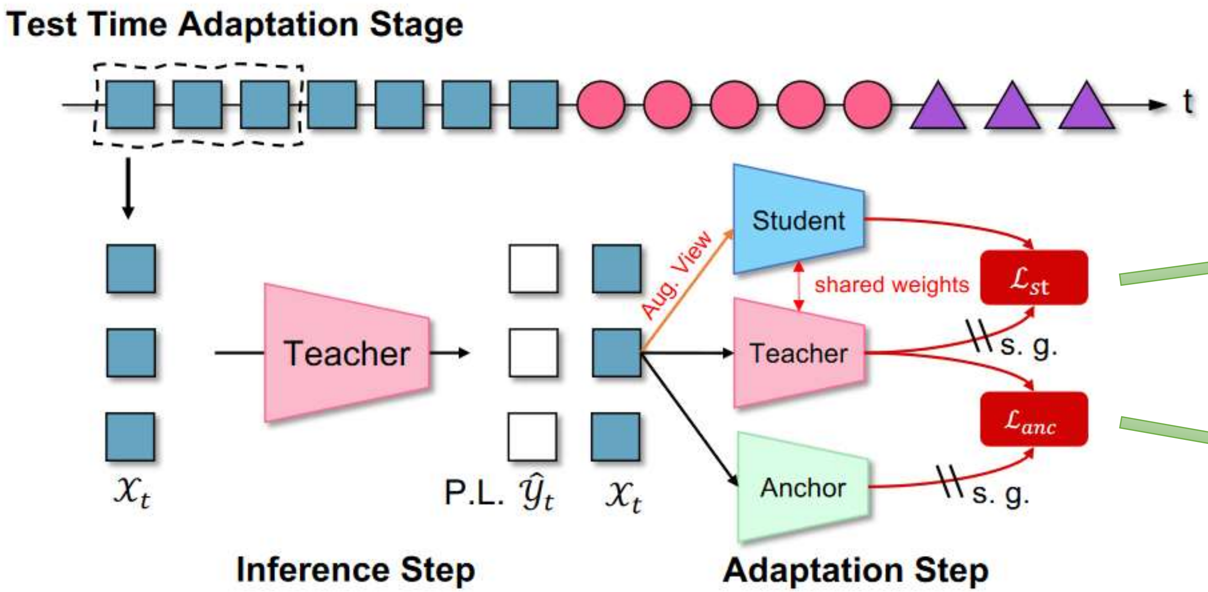
$$\sigma_k^{2\,t} = \sigma_k^{2\,t-1} + (1-\gamma)\left\{ -\delta_k^2 + \eta \sum_{b=1}^{B} \mathbb{1}(\hat{y}_b = k) \frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} \left[(F_{bhw} - \mu_k^{t-1})^2 - \sigma_k^{2\,t-1} \right] \right\} + \tag{3}$$

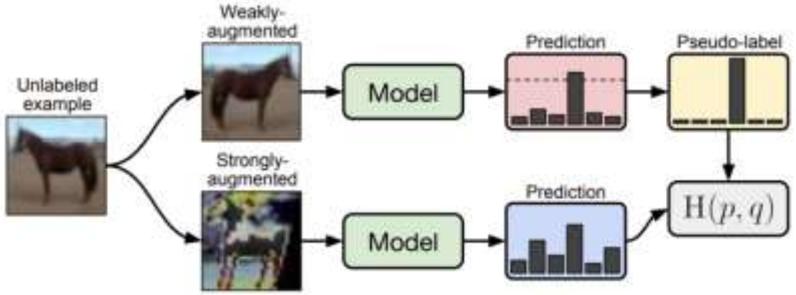
$$\gamma \cdot \frac{1}{K_c} \sum_{k'=1}^{K_c} \left\{ -\delta_{k'}^2 + \eta \sum_{b=1}^{B} \mathbb{1}(\hat{y}_b = k') \frac{1}{HW} \sum_{h=1}^{H}\sum_{w=1}^{W} \left[(F_{bhw} - \mu_{k'}^{t-1})^2 - \sigma_k^{2\,t-1} \right] \right\}.$$



Methods

Tri-Net Self-Training





$$\mathcal{L}_{st} = \frac{\sum_{b=1}^{B} \mathbb{1}(\mathcal{H}(\mathbf{p}_b^t) < H_0 \cdot \log K_c) \cdot \mathcal{H}(\hat{\mathbf{p}}_b^t, \mathbf{p}_b^s)}{\sum_{b=1}^{B} \mathbb{1}(\mathcal{H}(\mathbf{p}_b^t) < H_0 \cdot \log K_c)},$$

$$\mathcal{L}_{anc} = \frac{\sum_{b=1}^{B} \mathbb{1}(\mathcal{H}(\mathbf{p}_b^t) < H_0 \cdot \log K_c) ||\mathbf{p}_b^t - \mathbf{p}_b^a||_2^2}{K_c \sum_{b=1}^{B} \mathbb{1}(\mathcal{H}(\mathbf{p}_b^t) < H_0 \cdot \log K_c)}$$

Experiments

Method	Fixed Global Class Distribution (GLI-TTA-F)			
	$I.F. = 1$	$I.F. = 10$	$I.F. = 100$	$I.F. = 200$
TEST	43.50 / 43.50	42.64 / 43.79	41.71 / 43.63	41.69 / 43.47
BN [28]	75.20 / 75.20	70.77 / 66.77	70.00 / 50.72	70.13 / 47.34
PL [21]	82.90 / 82.90	72.43 / 70.59	70.09 / 55.29	70.38 / 49.86
TENT [40]	86.00 / 86.00	78.15 / 74.90	71.10 / 58.59	69.15 / 53.37
LAME [2]	39.50 / 39.50	38.45 / 40.07	37.48 / 41.80	37.52 / 42.59
COTTA [43]	83.20 / 83.20	73.64 / 71.48	71.32 / 56.44	70.78 / 49.98
NOTE [10]	31.10 / 31.10	36.79 / 30.22	42.59 / 30.75	45.45 / 31.17
TTAC [33]	23.01 / 23.01	31.20 / 29.11	43.40 / 37.37	46.27 / 38.75
PETAL [3]	81.05 / 81.05	73.97 / 71.64	71.14 / 56.11	71.05 / 50.57
RoTTA [46]	25.20 / 25.20	27.41 / 26.31	30.50 / 29.08	32.45 / 30.04
TRIBE	**16.14**(+6.86) / **16.14**(+6.86)	**20.98**(+6.43) / **22.49**(+3.82)	**19.53**(+10.97) / **24.66**(+4.42)	**19.16**(+13.29) / **24.00**(+6.04)

Method	Time-Varying Global Class Distribution (GLI-TTA-V)			
	$I.F. = 1$	$I.F. = 10$	$I.F. = 100$	$I.F. = 200$
TEST	43.50 / 43.50	41.95 / 43.65	40.74 / 43.83	40.53 / 43.77
BN [28]	75.20 / 75.20	71.36 / 67.70	70.35 / 53.07	70.88 / 50.67
PL [21]	82.90 / 82.90	74.74 / 72.12	73.03 / 57.53	72.49 / 54.20
TENT [40]	86.00 / 86.00	77.69 / 74.23	72.99 / 58.65	73.45 / 54.96
LAME [2]	39.50 / 39.50	38.02 / 40.15	36.51 / 42.16	36.24 / 42.16
COTTA [43]	83.20 / 83.20	75.29 / 71.87	73.83 / 56.80	74.97 / 56.47
NOTE [10]	31.10 / 31.10	29.52 / 29.23	30.02 / 29.88	29.71 / 30.28
TTAC [33]	23.01 / 23.01	32.25 / 32.12	36.84 / 37.13	37.96 / 38.07
PETAL [3]	81.05 / 81.05	75.19 / 71.65	72.71 / 55.73	73.76 / 53.51
RoTTA [46]	25.20 / 25.20	27.61 / 26.35	32.16 / 29.32	33.34 / 31.35
TRIBE	**16.14**(+6.86) / **16.14**(+6.86)	**20.92**(+6.69) / **22.40**(+3.95)	**22.44**(+7.58) / **25.50**(+3.82)	**23.10**(+6.61) / **27.03**(+3.25)

Experiments

Time	*t* →															
Method	motion	snow	fog	shot	defocus	contrast	zoom	brightness	frost	elastic	glass	gaussian	pixelate	jpeg	impulse	Avg.
TEST	85.15	83.45	75.88	97.09	81.68	94.52	77.93	41.23	77.07	82.48	89.73	97.81	79.31	68.50	98.17	82.00
BN [28]	73.64	66.07	52.81	84.49	85.05	82.66	61.96	36.04	68.60	56.44	84.85	85.31	52.29	60.77	85.05	69.07
PL [21]	66.55	60.43	49.46	**76.57**	79.23	81.04	65.35	51.48	75.62	69.74	89.04	92.36	86.84	92.09	97.83	75.58
TENT [40]	**64.37**	59.73	51.20	77.47	81.70	88.72	82.38	76.91	93.64	95.43	98.80	98.98	98.39	98.90	99.40	84.40
LAME [2]	85.93	84.57	77.29	97.47	81.92	94.72	78.41	41.49	77.67	84.07	90.25	98.21	79.61	68.64	98.76	82.60
EATA [29]	73.15	65.41	52.51	84.27	85.09	82.85	61.52	35.15	68.26	56.30	84.43	84.95	51.63	60.85	85.05	68.76
NOTE [10]	82.97	78.29	73.43	93.92	96.35	89.73	93.18	84.57	92.82	94.54	98.50	98.88	98.17	97.55	98.78	91.44
ROTTA [46]	74.86	70.02	55.26	85.55	85.37	78.61	61.00	**34.31**	64.65	52.83	76.16	85.43	**48.70**	52.41	78.37	66.90
TRIBE	69.52	**59.55**	**48.35**	79.27	**78.47**	**75.54**	**56.62**	35.19	**60.39**	**49.26**	**74.54**	**74.10**	50.08	**51.24**	**72.59**	**62.32**(+4.58)

Method	EMA Model	BatchNorm	Self-Training	Anchored Loss	CIFAR10-C	CIFAR100-C	Avg.
TEST	–	BN	–	–	41.71 / 43.63	47.53 / 45.91	44.62 / 44.77
ROTTA [46]	✓	Robust BN	✓	–	30.50 / 29.08	45.68 / 42.04	38.09 / 35.56
–	–	Robust BN	–	–	43.48 / 32.29	40.45 / 36.94	41.97 / 34.62
–	–	Balanced BN	–	–	29.00 / 26.38	39.55 / 36.59	34.28 / 31.49
–	–	BN	✓	–	37.67 / 38.94	37.12 / 44.77	37.40 / 41.86
–	–	Balanced BN	✓	–	36.58 / 65.88	37.21 / 44.83	36.90 / 55.36
–	–	BN	✓	✓	36.76 / 29.19	36.16 / 36.26	36.46 / 32.73
MT*	✓	Balanced BN	✓	–	23.76 / 25.18	36.01 / 35.72	29.89 / 30.45
TRIBE	–	Balanced BN	✓	✓	**19.53 / 24.66**	**32.31 / 34.98**	**25.92 / 29.82**

Experiments

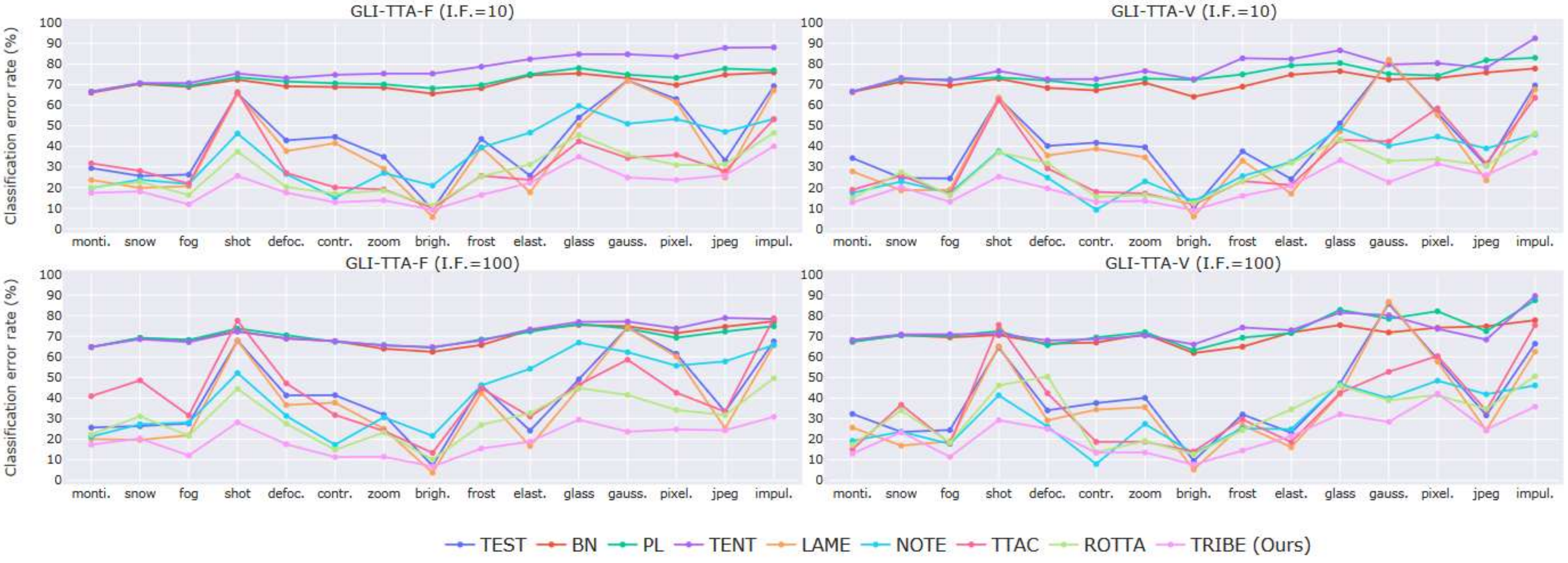


Figure 4: Performances on each individual domain (corruption) under GLI-TTA protocols on CIFAR10-C dataset.

Thanks